Author's response to reviews

Title: Rasch Fit Statistics and Sample Size Considerations for Polytomous Data

Authors:

Adam B Smith (a.b.smith@leeds.ac.uk)
Robert Rush (RRush@qmu.ac.uk)
Lesley J Fallowfield (l.j.fallowfield@sussex.ac.uk)
Galina Velikova (g.velikova@leeds.ac.uk)
Michael Sharpe (michael.sharpe@ed.ac.uk)

Version: 4 Date: 6 May 2008

Author's response to reviews: see over
Dear Ms. Puebla,

**Re: Rasch Fit Statistics and Sample Size Considerations for Dichotomous and Polytomous Data**

We are grateful for the opportunity to submit a revision of our manuscript for consideration for publication in *BMC Medical Research Methodology*.

We believe that we have addressed the concerns raised by the Editorial Board reviewer. In particular, we have emphasised in the revised manuscript that we are referring to putative Type I error rates as the “true” error rate is not known due to the real data (although the use of derived data is common practice in this field, e.g. Smith R & Suh, 2003) and have discussed this extensively in the limitations of the study.

The discussion of Type II error rates was introduced after comments made by Reviewer 2 in the first draft of the manuscript and it was never an aim of the study to explore these error rates, however we felt this was an important point and therefore included a discussion of this issue.

We are also very grateful to Prof. Bond’s (Reviewer 2) for his positive endorsement of our manuscript and the revisions made.

Yours sincerely,

Adam Smith, PhD
Corresponding author
On behalf of the co-authors
Reviewer 2

1. Comment 1:
   - pls use 'fit TO the model'

Reply to Comment 1:
   - The manuscript has been amended accordingly.

2. Comment 2:
   - .00 entries to the tables are unnecessary

Reply to Comment 2:
   - The Tables have been amended accordingly.
Comments by Editorial Board reviewer:

Comment 1:

Technically, I think both reviewers are right, although they are evaluating different aspects of the paper. Reviewer 1 is correct in that an evaluation of type I error is not possible without either a theoretical or simulation approach. In the latter, the "truth" is generated and the ability of the various approaches to capture the truth is estimated. The authors are showing how results vary with sample size. If they base their conclusions strictly on those patterns, I think the paper still makes a useful contribution. When they try to extend their conclusions to a discussion of type I error, I agree with Reviewer 1 that such an interpretation is problematic.

Reply to Comment 1:

The intention of this manuscript was to demonstrate the association between sample size and the variation in the various fit statistics, in particular the t-statistics. We have emphasised in the conclusion that we are using the term Type I error to refer to misidentification of misfitting items in this study, and have acknowledged and re-emphasised in our discussion of the limitations that simulated data is best suited for the evaluation of the "true" error rates. However, we would also like to point that the use of "real" data is quite common in this field (e.g. Smith & Suh, J Applied Measurement, 2003).

For instance, “The results of this study demonstrated that Type I error rates – defined strictly in this study as falsely rejecting an item as not fitting the Rasch model - for the t-statistic were at least twice those of the corresponding fit statistic for both infit and outfit for both Rasch Models” (page 11).

“Further analysis of the individual item fit and sample size suggested that although in the majority of cases there was agreement between mean square and t-statistics in terms of identifying fit and misfit (>50% for both models and questionnaires), there were discrepancies in Type I error rate as defined in this study and a lack of sample size invariance for the t-statistics.” (page 12).

“The results of the study suggest that the t-statistics are highly dependent on sample size which has the effect of inflating putative Type I error rates.” (page 12)

“However, we acknowledge that estimates of true Type I error rates are more optimally derived from simulated data where fit and misfit may be artificially manipulated” (page 13).

Comment 2:

The problem is that we don't really know whether the association of the t-test results with sample size indicates that type I error is too high at larger sample sizes, or type II error is too high at smaller sample sizes. I was a bit confused by the authors' reply in this regard. I wasn't sure which of these two conclusions they were really aiming for. Again, though, if the message is simply to show the association with sample size, then the paper does contribute something. One could debate the importance of such a contribution, but it seems to me to be valuable enough to warrant publication.

Reply to comment 2:

We included a discussion on Type II errors following comments on the original draft by Reviewer 2. As we stated in our reply it was never our intention to explore Type II errors in this study, however previous research we carried out allowed us to discuss error rates and misidentification of misfit and the potential impact on the clinical utility of the instruments.